Journal of Applied Psychology
1989, Vol. 74, No. 2, 347-355

0021-9010/89/$00.75

Psychopathy, Threat, and Polygraph Test Accuracy

Christopher J. Patrick
University of British Columbia

William G. Iacono
University of Minnesota

The accuracy of the control question polygraph test with psychopaths was evaluated within a realistically threatening context. Subjects were 24 psychopathic and 24 nonpsychopathic male prison inmates. Within each diagnostic group, equal numbers of "guilty" and "innocent" subjects were tested by experienced professional polygraphers regarding their involvement in a mock theft by using standard control question procedures. A group contingency threat was devised in which subjects believed that their personal performance could decide benefits or penalties for the sample as a whole. Guilty psychopaths were detected just as easily as guilty nonpsychopaths, and the majority of guilty subjects (87%, excluding inconclusives) were correctly identified. However, innocent subjects were identified with only 56% accuracy, and an analysis of false positive errors suggested that the subjective impact of the threat was a critical factor in these outcomes.

An important issue concerning the use of the polygraph ("lie detector") test in criminal investigations is its accuracy with psychopathic individuals, who are reputed to be skilled liars and manipulators with a limited capacity for anxiety or guilt (cf. Cleckley, 1976). The possibility that psychopaths might be able to beat the lie detector is a concern, particularly in view of recent research (Hare, 1981) that suggests that psychopaths constitute a sizable proportion of criminal offenders, particularly violent and repetitive offenders. Laboratory data showing that psychopaths tend to be nonreactive to unpleasant stimuli (Hare, 1978) lend credence to the idea that psychopaths might be less disturbed by the relevant questions on a standard polygraph test and therefore less likely to produce clear deceptive outcomes. Also suggestive are the results of a study by Waid, Orne, and Wilson (1979), which found that subjects scoring low on a personality scale measuring socialization, a construct related to psychopathy, were more likely to escape polygraphic detection.

However, the one study to date that has examined this issue directly (Raskin & Hare, 1978) concluded that imprisoned psychopaths who committed a mock theft could be easily detected by using the control question test (CQT), the polygraph procedure normally used in police investigations. This study is noteworthy in that it used rigorous diagnostic procedures to identify psychopaths, and polygraph tests were conducted by an experienced examiner, using field procedures. Excluding inconclusive test results, the overall hit rate for psychopathic and nonpsychopathic prisoners, of whom half were "guilty" and half "innocent," was 95.5% (with 0% of guilty subjects and only 9.5% of innocent subjects misclassified).

Because of their important social policy implications, the Raskin and Hare (1978) findings have been the subject of intense controversy. In particular, Lykken (1978, 1981) has challenged the study's credibility on several grounds. In the first place, according to Lykken, the hit rate in the Raskin and Hare study was likely inflated because Raskin, who administered the tests, also scored the charts and hence was in a position to use extrapolygraphic cues (e.g., subject behaviors and comments) to make judgments of guilt and innocence. In support of this argument, Lykken (1978) noted that the hit rate for Raskin and Hare's psychopaths was appreciably higher than that obtained in studies using nonpsychopaths and blind scoring procedures. A second objection was that the mock crime paradigm used by Raskin and Hare provides little indication of how subjects (psychopathic and nonpsychopathic) would fare under the manifestly more complex and stressful circumstances of a real-life polygraph test. Appraisers of the CQT have pointed out that the technique is based on unproven assumptions, and concern exists that it may yield a high rate of incorrect decisions in real-life situations where, in contrast to most laboratory paradigms, the outcome of a test may have serious consequences (see Iacono & Patrick, 1987, 1988, for a discussion of CQT theory and validity issues). A third point was that subjects in the Raskin and Hare study were responding purely on the basis of a reward incentive (i.e., they were told that they could keep the "stolen" $20 if they beat the test), a context wherein psychopaths have been shown to exhibit as much arousal as nonpsychopaths

This research was supported in part by Grant No. 410-84-0614 from the Social Sciences and Humanities Research Council of Canada.

This manuscript was based on a dissertation submitted by Christopher J. Patrick (who is now at the University of Florida in Gainesville) in partial fulfillment of the requirements for the degree of doctor of philosophy at the University of British Columbia. Preliminary findings were presented at the 94th Annual Convention of the American Psychological Association, Washington, DC, August 1986, and the 26th annual meeting of the Society for Psychophysiological Research in Montreal, Quebec, October 1986.

We are grateful to the staff and inmates of the Lower Mainland Regional Correctional Center, Burnaby, British Columbia, for their cooperation and assistance, and to Harold Stevenson and John Weller for conducting the polygraph examinations. We would also like to thank Robert Hare, David Lykken, and David Raskin for their helpful comments during the planning stages of this project.

Correspondence concerning this article should be addressed to William G. Iacono, Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

(Schmauk, 1970). According to Lykken, psychopath/nonpsychopath detectability differences would be expected only in situations where subjects are motivated by fear or guilt. Likewise, Hare (1978) has expressed the view that autonomic differences between psychopaths and others are task-specific, occurring more readily in situations that are monotonous or threatening than in those that are interesting or exciting.

This study was designed to reevaluate the accuracy of the CQT with psychopathic individuals. A constructive replication (Lykken, 1968) of the Raskin and Hare study was carried out in which subjects were tested regarding a mock theft under conditions of threat rather than reward. The tests were scored by the original examiner and by a separate evaluator who examined only the chart data, to determine whether access to extrapolygraphic information produced a difference in overall accuracy. In addition, physiological activity during the test was recorded simultaneously on laboratory and field model polygraphs to provide us with quantitative physiological data in addition to basic field measures.

Method

Subjects

Subjects were 24 psychopathic and 24 nonpsychopathic male prisoners (mean ages 23.83 and 27.96 years, respectively) recruited from a provincial prison in Burnaby, British Columbia. Diagnostic classifications were made by using Hare's (1980) psychopathy checklist, an instrument whose psychometric properties and relevance to Cleckley's (1976) conception of psychopathy are well-established (Hare, 1980; Schroeder, Schroeder, & Hare, 1983). Inmates scoring in the lowest third of the sample were classified as nonpsychopaths, and those scoring in the highest third were classified as psychopaths. To further ensure diagnostic purity, all psychopathic subjects were also required to meet the DSM-III (American Psychiatric Association, 1980) criteria for antisocial personality disorder (APD). Both checklist and APD assessments were based on lengthy (1½–2 hour) structured interviews supplemented by case file information, which included reports by correctional officers and medical/psychiatric personnel, together with criminal records, police and inmate versions of the offences, and daily progress summaries describing institutional behavior. Interview and case file data for each inmate were reviewed independently by two trained diagnosticians, and only inmates on whom there was a consensus opinion were included for study. Checklist scores from the two diagnosticians were averaged to obtain a composite score for each subject.

A total of 107 inmates were assessed in order to obtain the final study sample of 48. The mean checklist score for the sample as a whole was 27.50 (SD = 7.47). The mean checklist score for subjects classified as nonpsychopathic was 19.36 (SD = 4.79; range = 9.5 to 26.0), and for psychopathic subjects it was 35.95 (SD = 1.80; range = 33.5 to 41.5), closely matching score distribution figures reported in previous studies by Hare (1980, 1985). The product–moment correlation between the checklist scores produced by the two diagnostic raters was .93, indicating a high level of interrater reliability. A kappa coefficient (Cohen, 1960) of .82 was obtained between the two sets of APD diagnoses, indicating a high degree of interrater agreement that cannot be attributed to chance.

Of the 59 inmates who were assessed but were not part of the final study sample, 38 were excluded because they failed to meet the diagnostic criteria described. An additional 14 inmates were either discharged from the prison or withdrew from the study before they could be tested. A final group of 7 inmates participated in the lie detection phase but were excluded for reasons that will be discussed in the next section.

Crime and Motivational Manipulations

In response to Lykken's (1978, 1981) comments, an effort was made to create a realistic threat context for the polygraph tests. The threat was founded on the peculiar pressures that exist within the inmate subculture, as sociological studies have documented (see, e.g., Clemmer, 1960; Sykes, 1958; Thomas & Peterson, 1977). In brief, there exists an unwritten code among prisoners that places a premium on strength, courage, and solidarity. A failure to live up to group expectations can provoke responses much stronger than mere disapproval: Peer labeling in the prison environment frequently leads to ostracism, persecution, and physical brutality. Because of this unusually strong peer pressure atmosphere, we anticipated that study participants would be quite concerned about their polygraph test outcomes if they believed that their poor performance could cause their fellow inmates to be penalized.

Potential subjects were offered $2 to participate in the initial assessment phase of the study, plus an additional $10 if they were chosen to participate in the "lie detection experiment." The latter payment was made without regard to polygraph test outcomes. Instead of offering individual rewards for truthful polygraph test results, monetary bonuses for the volunteer group as a whole were made to depend upon each individual's performance. Potential volunteers were advised that we were interested in the ability of inmates "to beat a lie detector test when there is something important at stake." They were told that participants in the lie detection experiment would be tested regarding their involvement in a mock theft (of which half would be "guilty"), and that on completion of the study each participant would receive a $20 bonus if the following condition was satisfied: that not more than 10 of the 48 participants, whether guilty or innocent of the crime under investigation, failed the polygraph test. Twenty dollars represented a considerable sum of money to these individuals, because they earned as little as $1 per day at their prison jobs and most had no other sources of income. As a further means of increasing their concern, subjects were led to believe that if more than 10 participants were judged to be deceptive and the group lost the bonuses, all participants would learn who was responsible because they would be provided with a list at the end of the study indicating who had passed and who had failed the test. Participants were told that no information about subject test performance would be released before the end of the study, under the rationale that knowledge of other prisoners' performance might affect a subject's confidence and his ability to beat the test. However, to eliminate any possibility of harm, the bonuses were ultimately awarded without regard to the group's performance, and subjects were advised privately of their test outcomes.

The lie detection phase of the study was conducted in the health care center of the prison. Half of the psychopaths and half of the nonpsychopaths were assigned to the guilty condition, the other half of each group to the innocent condition. If guilty, the subject was instructed to sneak down the hall from the testing room and steal $20 from the pocket of a jacket in the doctor's office, a room normally off-limits to prisoners. He was instructed to be quick and to avoid being seen, and to hide the $20 in his pocket until after the polygraph test. (In terms of the arousal value of the "theft," it is worth noting that inmates are normally forbidden to possess money inside the prison.) Subjects assigned to the innocent group were made aware of the general details of the crime committed by guilty subjects, but remained in the testing room at all times. All participants were carefully instructed to deny having committed the mock crime during the polygraph test and to avoid making statements

that might reveal their criterion status.[1] The experimenter stressed to each subject the fact that his peers were depending on him to obtain a truthful test result.

Of the total of 55 subjects who participated in this phase of the study, 7 were not used. Six of these (three psychopaths and three nonpsychopaths) were excluded either for procedural reasons (i.e., initial problems with the instructions; irregularities in the commission of the mock theft) or because of medical conditions that raised doubts about their suitability for testing. The remaining subject, a nonpsychopath, withdrew from the study after the polygraph pretest interview, explaining, "I'm too nervous . . . I don't want to let the other guys down." It is also worth noting that 5 other study volunteers who had cooperated with the assessment phase and expressed strong interest in taking the polygraph test subsequently withdrew when asked to appear for the lie detection experiment. Of the 55 subjects who did appear, 24 (12 psychopaths and 12 nonpsychopaths) made unsolicited remarks either before or after the test acknowledging their fear or expressing pessimism about their performance.

Apparatus

The polygraph examinations were conducted in a large medical testing room (approximately 8 m × 5 m) situated in a quiet wing of the prison's health care center. The examiner conducted his test by using a five-channel Lafayette Model 761-99G field model polygraph. Thoracic and abdominal respiration measures were obtained from pneumatic tubes positioned around the subject's upper thorax and abdomen. Skin resistance response (SRR) recordings were obtained from two stainless steel electrodes attached to the volar surface of the first and third fingers of the subject's right hand. Mechanical and electronically enhanced cardiovascular tracings were obtained from a pneumatic pressure cuff positioned around the upper portion of the subject's left arm.

The subject's physiological responses during the polygraph test were also recorded on a 4-channel Beckman Type R612 Dynograph situated in an adjoining storage room that also contained video equipment to record the session. An experimenter (the first author) monitored the laboratory polygraph and video equipment while the polygraph examiners conducted tests using the field instrument. The Beckman polygraph recordings included skin conductance (SC), which was measured from 1-cm Ag/AgCl electrodes filled with a Unibase/saline paste and placed on the palm of the right hand, and heart rate (HR) activity, which was monitored from the subject's right wrist and left ankle (EKG Lead II).

Procedure

On arrival at the prison's health care center, each inmate subject completed the Socialization (So) scale of the California Psychological Inventory (Gough, 1969). Although findings reviewed by Hare (1978) had indicated that clinically diagnosed psychopaths who obtain low So scores might form a purer group of primary psychopaths, the findings of Raskin and Hare (1978) failed to support this notion. The availability of So scale data in our study permitted us to readdress this question.

Following the experimenter's instructions, provision of written consent, and commission of the mock theft in the case of guilty subjects, suspects were tested for complicity by one of two polygraphers. Both examiners met the following qualifications: (a) formal training at an established American polygraph school; (b) membership in the major North American professional associations for polygraph examiners (including the American Polygraph Association); (c) more than 25 years of police experience; and (d) over 10 years' experience conducting field polygraph tests with criminal suspects. Subjects were assigned randomly to the two polygraphers with the proviso that each examined equivalent numbers of guilty and innocent psychopaths and nonpsychopaths. The polygraphers were blind to guilt and psychopathy status on each examination, and no information was provided regarding the base rate of guilt in the sample. In addition, feedback on the accuracy of decision making was withheld until all subjects in the study had been tested.

The polygraph examination consisted of a pretest interview followed by a test phase in which the examiner recorded physiological responses to test questions. Although the same general procedure was followed by both examiners, each used the testing style and examination forms common to his own practice. The pretest interview was devoted to the collection of biographical information from the subject, a discussion of the incident under investigation, a review of the polygraph test questions, and an explanation of the physiological measures to be obtained and the theory underlying polygraphic lie detection (cf. Raskin & Hare, 1978).

The polygraph test was a version of the control question test (CQT) known as the modified zone of comparison technique, in which relevant and control questions appear in pairs. Because the main issue was the same, the question sequence used in this study paralleled closely the sequence used by Raskin and Hare (1978), and was as follows:

1. (Irrelevant) Is your true name _____ ?
2. (Sacrifice relevant) Regarding whether you took that $20 from the pocket of a jacket in the Doctor's office today, do you intend to answer truthfully each question about that?
3. (Outside issue) Are you completely convinced that I will not ask you a question on this test that has not already been reviewed?
4. (Control) During the first _____ years of your life, do you now remember ever stealing any money?
5. (Relevant) Did you take that $20?
6. (Control) During the first _____ years of your life, do you now remember ever stealing anything from someone who trusted you?
7. (Relevant) From a jacket in the Doctor's office today, did you take that $20?
8. (Outside issue) Is there something else you are afraid I will ask you a question about, even though I told you I would not?
9. (Control) During the first _____ years of your life, do you now remember ever taking anything of value from an employer?
10. (Relevant) Do you know where that $20 is at this time?

Each of the relevant questions dealt with the mock theft issue, whereas the control questions covered related issues from the subject's past (see Barland & Raskin, 1975, for a discussion of the function of the other CQT items). All of the test questions were identical for each subject except the control questions, which were phrased to include all of the subject's life up to his last birthday. The examiner reviewed each of these items with the subject and adjusted their wording in response to admissions by the subject (cf. Raskin & Hare, 1978) so that the subject was able to provide a *no* response to each of the relevant and control questions.

---

[1] It became evident after testing our first two subjects that without very clear instructions on how to deny involvement in the mock theft, including a rehearsal of answers to crime-relevant questions, some individuals would make obvious verbal blunders and give away their crite rion status. One subject appeared to become nervous and confused when the examiner asked him "Regarding the theft of $20 . . . , do you intend to answer truthfully each question about that?" and replied, "We're supposed to fool you, right?" Another, when asked "Do you know where that $20 is at this time?" responded "Yes." Without precautions similar to those we took, it is conceivable that subjects' comments during the polygraph test could have strongly influenced the examiner's opinions in the Raskin and Hare (1978) study.

Following the pretest interview, the examiner placed attachments for both the Lafayette field polygraph and the Beckman laboratory polygraph on the subject and completed two further procedures before beginning the actual polygraph test. The first was to provide a written, categorical judgment reflecting his belief in the subject's guilt or innocence on the basis of his pretest interactions with the subject, together with a rating on a 7-point scale reflecting his confidence in that belief (1 = *no confidence*, 7 = *complete confidence*). The purpose of this procedure was to permit an analysis of each examiner's ability to detect guilt accurately when using pretest behavioral cues. In addition, the examiner also conducted a demonstration or stimulation test (cf. Barland & Raskin, 1975; Reid & Inbau, 1977) to provide the subject with a graphic demonstration of the polygraph's ability to distinguish between a truth and a lie.

The actual polygraph test consisted of a minimum of three charts, with additional charts obtained if the results were not obvious after the third question set. Following standard field procedure, the positions of the control items in the question sequence were reordered for the second chart, and the positions of the relevant items were reordered for the third. After each chart, the examiner asked the subject whether any of the questions had bothered him, with specific reference to the control questions. These items were further modified or clarified in response to new admissions from the subject. In contrast, the wording of the relevant questions was never changed.

On completion of the polygraph test, the examiner departed and the experimenter returned from the adjacent room. Prior to debriefing, the subject was asked to complete the State form of the State–Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970), with the instructions modified to make reference to the subject's state during the polygraph test.

Quantification of Physiological Measures

Lafayette field polygraph. After each polygraph examination, the original examiner conducted an on-the-spot numerical analysis by using the scoring criteria he normally used in his field practice. At a later date, the other study examiner rescored each set of field charts, which yielded an independent blind analysis. The numerical analysis criteria used by the two examiners were similar to those used by Raskin and Hare (1978). On the basis of the relative magnitude of responses to test questions, a score of +3 to −3 was assigned to each control-relevant question pair for each of the three physiological channels (respiration, skin resistance, and cardiovascular), with a positive score reflecting greater response to the control question and a negative score reflecting greater response to the relevant question. A total score for the first three polygraph charts was obtained by summing these scores over question pairs, channels, and charts. Subjects with total scores of +6 or higher were classified as truthful, those with scores of −6 or lower were classified as deceptive, and those with scores between +5 and −5 were classified as inconclusive. In cases where the examiner was unsure after three charts whether his subsequent numerical analysis would yield a conclusive score, an additional chart was administered. Although a fourth chart was obtained in 12 cases, no more than three charts were ever required by the original examiner to reach a conclusive score.

Beckman laboratory polygraph. The physiological responses recorded on the laboratory polygraph were quantified objectively by trained technicians who were blind to the group membership of the subjects involved. The following two global measures of physiological arousal during the polygraph test were derived:

Skin conductance level (SCL): The mean SCL (in micromhos) for each chart was calculated on the basis of prestimulus levels for all 10 questions.

Heart rate level (HRL): HRL (in beats per minute) was defined as the mean HR for the last three beats preceding the beginning of the question, and a mean HRL for each chart was calculated on the basis of the 3-beat prestimulus levels for all 10 questions.

Table 1
Decisions Based on Numerical Scores: Original and Blind Chart Analyses

Analysis and group	Decision		
	Deceptive	Truthful	Inconclusive
Original			
Guilty			
Psychopath	11	1	0
Nonpsychopath	11	1	0
Innocent			
Psychopath	5	7	0
Nonpsychopath	5	7	0
Blind			
Guilty			
Psychopath	10	2	0
Nonpsychopath	10	1	1
Innocent			
Psychopath	3	5	4
Nonpsychopath	5	5	2

Results

Self-Reported Anxiety

A two-way (Criterion Group × Psychopathy Diagnosis) analysis of variance was performed to examine group differences in state anxiety associated with the polygraph test, as measured by the STAI scale (where higher scores reflect greater anxiety). Neither of the main effects was significant, nor was the interaction, suggesting similar levels of self-reported anxiety for guilty and innocent psychopaths and nonpsychopaths. The mean STAI score for the sample as a whole was 44.0 (SD = 9.23), with individual scores ranging from 27 to 62.

Numerical Evaluations of Field Polygraph Charts

Accuracy of field decisions. Using an inconclusive region of −5 to +5, subjects were classified according to their total scores for three charts, because the classification results were the same whether based on three charts or (where available) four. On the basis of the original analysis of the charts (see Table 1), there were 75% correct, 25% incorrect, and 0% inconclusive decisions, with the majority of errors (10/12) being false positives (i.e., innocent subjects classified as deceptive). The accuracy figures for psychopaths and nonpsychopaths were identical. To determine whether examiner styles contributed to differences in accuracy, the hit rates for the two original examiners (which were virtually identical) were compared by using separate chi-square analyses for guilty and innocent subjects. No significant differences emerged.

Table 1 also shows the classification results for guilty and innocent psychopaths and nonpsychopaths based on a blind analysis of the charts. The blind evaluations produced 62.5% cor-

rect, 22.9% incorrect, and 14.6% inconclusive decisions. Excluding inconclusives, the accuracy rate was 73.2%, and Yates-corrected chi-square tests comparing the classification figures for psychopaths and nonpsychopaths revealed no significant differences in either the guilty or the innocent condition.

Product–moment correlations between the original and blind total scores and individual channel scores (respiration, skin resistance, and cardiovascular) generally revealed a high degree of correspondence, suggesting that the original examiners relied primarily on the chart data in their analyses. The only relationship that failed to exceed .80 was the correlation between the respiration channel scores for the two chart analyses ($r = .55$); correlations for the total score, skin resistance score, and cardiovascular score variables were .85, .90, and .82, respectively. To examine the comparative accuracy of classifications made on the basis of original and blind scorings, we conducted a Yates-corrected chi-square test comparing hit rates for guilty and innocent subjects (excluding inconclusives) by using the data from the original examiner to set the expected frequencies for the blind examiner. The test was nonsignificant, $\chi^2(1, N = 41) = 0.57$. There were seven cases in which the original examiner's total score was conclusive when the blind examiner's was inconclusive. There was only one case (a guilty psychopath) in which the two analyses produced conclusive decisions that were different. From this point on, discussion of the field polygraph data is restricted to the blind results because (a) the original and blind results were very similar, and (b) the blind analysis can be assumed to reflect a more unbiased assessment of the chart tracings (cf. Lykken, 1981).

We used a Yates-corrected chi-square test to determine whether the overall accuracy of the blind scoring decisions (excluding inconclusives) exceeded a chance level of 50%. The test was significant, $\chi^2(1, N = 41) = 11.20, p < .01$. However, further chi-square analyses performed separately on guilty and innocent subjects revealed that although the decisions made on guilty subjects exceeded chance-level accuracy, $\chi^2(1, N = 23) = 11.14, p < .01$, the decisions made on innocent subjects did not, $\chi^2(1, N = 18) = .06$. Excluding inconclusives, the hit rate for guilty subjects in the present study was 87.0% (20/23); for innocent subjects it was 55.6% (10/18). A further question of interest was whether the hit rates for guilty and innocent subjects reported by Raskin and Hare (21/21 = 100% and 21/23 = 91.3%, respectively) differed significantly from those obtained in our study. Yates-corrected chi-square tests indicated that although the hit rates for guilty subjects in the two studies were not significantly different, the hit rate for innocent subjects in the Raskin and Hare study was significantly higher than in our study, $\chi^2(1, N = 41) = 5.19, p < .05$.

A final issue investigated was the degree to which alterations in the cutoffs of the inconclusive region might lead to improved hit rates (cf. Raskin & Hare, 1978). Figure 1 depicts the results in terms of percentage accuracy of decisions and percentage inconclusive results for guilty and innocent subjects by using cutoffs ranging from 0 to ±12 to classify the scores based on three charts. The optimal cutoff region was ±2, at which point 88% of the guilty subjects and 59% of the innocent subjects were classified correctly, with only 4% (2/48) inconclusives for the

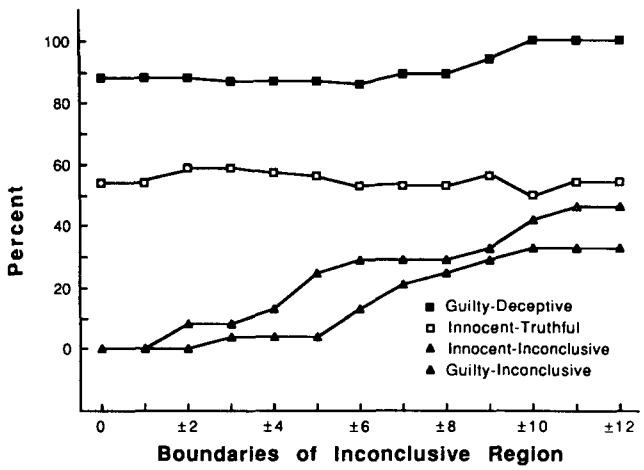

Figure 1. Accuracy of decisions and percentage of inconclusives for varying inconclusive ranges.

sample as a whole. Raskin and Hare also reported an optimal cutoff region of ±2, although in that study the accuracy of decisions for both guilty and innocent subjects exceeded 95%, using the ±2 cutoffs. The reason why this was not so in our study was obvious from the distributions of total blind numerical scores for guilty and innocent subjects: Whereas the scores for guilty subjects were strongly clustered toward the negative end of the continuum, the scores for innocent subjects were more evenly distributed over a range of −26 to +22.

Effectiveness of numerical scores. Separate three-way analyses of variance were used to compare the individual channel and total blind scores for guilty and innocent subjects falling into the four categories of psychopathy and socialization (cf. Raskin & Hare, 1978). Following the procedure used in the earlier study, the 12 subjects within each of the four cells defined by criterion group and psychopathy diagnosis were divided into low and high socialization subgroups on the basis of their So scale scores. Table 2 shows the mean channel and total scores for the groups involved in these analyses. For the total score variable, the only significant effect was the criterion group (guilty vs. innocent) difference, $F(1, 40) = 14.18, p < .01$. The criterion group difference also represented the only significant effect in the other three analyses, involving respiration score, $F(1, 40) = 8.76, p < .01$; skin resistance score, $F(1, 40) = 12.56$, $p < .01$; and cardiovascular score, $F(1, 40) = 6.70, p < .05$. There were no significant main effects for psychopathy diagnosis or socialization level, nor were any of the interactions significant.

To obtain a quantitative assessment of the discriminative value of the three physiological channels in terms of separating guilty and innocent subjects, we conducted a stepwise discriminant analysis. A significant discriminant function was obtained, $F(2, 45) = 8.39, p < .01$. The discriminating variables were skin resistance score and respiration score, with standardized discriminant function coefficients of 0.72 and 0.46, respectively. Cardiovascular score did not enter into the discriminant function because its potential independent contribution to the

Table 2
Mean Blind Total and Channel Numerical Scores for Charts 1–3

Group	Score			
	Respiration	SRR	Cardiovascular	Total
Psychopath				
Low So				
Guilty	−2.83	−8.50	−4.33	−15.67
Innocent	1.00	0.50	2.50	4.00
High So				
Guilty	−4.83	−6.00	0.00	−10.83
Innocent	0.00	−0.67	1.33	0.67
Nonpsychopath				
Low So				
Guilty	−2.67	−8.00	−5.33	−16.00
Innocent	0.33	−1.17	1.50	0.67
High So				
Guilty	−3.67	−6.67	−4.50	−14.83
Innocent	2.33	−1.00	−1.00	0.33
Combined				
Guilty	−3.50	−7.29	−3.54	−14.33
Innocent	0.92	−0.58	1.08	1.42

Note. SRR = skin resistance response; Low So and High So = low and high socialization subgroups, based on Socialization scale scores.

equation was minimal. Classifications based on the discriminant function scores were 72.9% accurate, which is almost identical to the overall hit rate of 73.5% obtained when an optimal cutoff region of ±2 was used to classify total numerical scores (see Figure 1).

Examiner Pretest Guilt Judgments

A comparison of the categorical pretest judgments (*deceptive* or *truthful*) made by the examiners with the actual criterion status of subjects showed that both examiners performed at precisely chance level, each making correct judgments in 50% of the cases they rated. There was no difference in the accuracy of pretest judgments for guilty versus innocent subjects or for psychopaths versus nonpsychopaths. A further question of interest was the degree to which the pretest guilt judgments agreed with decisions based on the original numerical analysis, irrespective of ground truth. Chi-square analyses indicated that the overall level of agreement was no better than chance for either category of chart-based decisions (*deceptive* or *truthful*), suggesting that in general the original chart analyses were not influenced by the examiners' subjective assessments of guilt.

False Positive Errors

We made an attempt to identify factors associated with false positive errors (i.e., innocent subjects misclassified as deceptive). One question of interest was whether the false positives differed systematically from the true positives (i.e., guilty subjects correctly classified) in terms of their polygraph test responses. A two-tailed t test between the total score means for the two groups (−16.38 and −18.35, respectively) was not significant, $t(26) = 0.63$. Moreover, a stepwise discriminant analysis using the separate channel scores (respiration, skin resistance, and cardiovascular) derived from the field polygraph charts to differentiate the groups failed to produce a significant discriminant function, suggesting that the true positive and false positive subjects displayed a similar pattern of reactions on the polygraph test.

A second question of interest was whether there were any unique subject characteristics associated with innocent subjects who produced deceptive test results (false positives) as compared with innocent subjects who produced truthful results (true negatives). We conducted a discriminant analysis by using the following subject variables as discriminators: age, education, self-reported anxiety during the polygraph test (STAI-State score), and level of socialization (So scale score). The analysis produced a significant discriminant function, $F(3, 14) = 4.60$, $p < .05$, with STAI score, age, and So score contributing to the equation. It appeared that So score contributed mainly as a suppressor variable, judging from its moderate correlations with STAI score and age (.45 and .31, respectively) and the fact that the So score means for the false positive and true negative groups were nearly identical (M = 24.00 and 23.50, respectively). On the other hand, group means for the STAI score and age variables were substantially different, with the STAI score difference being significant at the univariate level, $F(1, 16) = 6.21$, $p < .05$. The standardized discriminant function coefficients for STAI score and age were 1.03 and .75, respectively.

An inspection of the discriminant function indicated that subjects in the false positive group reported experiencing less anxiety during the polygraph test and tended to be younger than subjects in the true negative group. This finding seemed counterintuitive in light of arguments that false positive outcomes are more likely when fearfulness is high (cf. Lykken, 1974). However, informal observations during the experiment suggested that the younger inmates (a) appeared most concerned about the peer threat associated with the polygraph test (as suggested by pretest behaviors and a greater tendency to drop out between the assessment and test phases) and (b) tended to volunteer for the study more to prove themselves to their peers than for personal reasons, such as curiosity or greed. We hypothesized that false positive outcomes were related to a repressive coping style, in which attempts were made to conceal underlying fears behind a facade of stoicism.

To test this hypothesis, we conducted a second stepwise discriminant analysis in which two gross measures of physiological arousal during the polygraph test, tonic skin conductance and tonic heart rate (i.e., mean prestimulus levels recorded by the Beckman polygraph over Charts 1–3), were included as discriminators along with STAI score and age. A fifth discriminator, psychopathy diagnosis (1 = *psychopath;* 2 = *nonpsychopath*), was also included because of its possible relation to the fear dimension. The resulting discriminant function was highly significant, $F(5, 12) = 9.39$, $p < .01$, with all five variables contributing to the equation. As predicted, false positive outcomes tended to be associated with (a) a diagnosis of *nonpsychopathic,* (b) lower age and STAI scores, and (c) higher tonic HR and SC scores (see Table 3). Actually, psychopathy diagnosis contributed more as a suppressor variable than as a direct discrimina-

Table 3
False Positives Versus True Negatives: Group Means for Discriminating Variables

Variable	Group	
	False positives[a]	True negatives[b]
STAI score	35.87	45.70
Psychopathy diagnosis[c]	62.50	50.00
Age	22.25	28.60
Tonic HR	84.29	79.34
Tonic SC	10.38	7.92

Note. STAI = State–Trait Anxiety Inventory; HR = heart rate; SC = skin conductance.
[a] $n = 8$. [b] $n = 10$. [c] Tabled values for psychopathy diagnosis variable represent percent of nonpsychopaths in each group.

tor because of its moderate correlations with the STAI and tonic SC variables (i.e., the innocent psychopaths generally produced lower STAI scores and higher tonic SC scores than the innocent nonpsychopaths, regardless of test outcome). Classifications of false positive versus true negative cases based on the discriminant function scores were 94.4% accurate, with the scores for the two groups forming two discrete, minimally overlapping distributions.

Discussion

As described earlier, one of the major criticisms of the Raskin and Hare (1978) study was that the subjects were not tested under conditions of threat, thereby compromising the external validity of the findings. To address this issue, participants in our study were led to believe that their poor performance could cause their fellow inmates to be deprived of a substantial (by prison standards) sum of money and that those responsible would be identified to the group. Because of the strong peer pressure atmosphere within the prison and the potential for scapegoating and physical aggression, we expected that the threat of disclosure would provide a strong incentive for subjects to pass the polygraph test. Informal indexes of fear, including subject attrition and participants' remarks to the experimenter during the test phase of the study (e.g., "I can feel my heart racing"; "I hope they [the other inmates] don't beat my head in if I fail") suggested that our subjects were indeed concerned about their test outcomes. More objective evidence that subjects were concerned about their performance was derived from the anxiety scale and tonic HR data. The mean STAI score of 44.0 produced by subjects in the present study was significantly higher than the mean of 37.1 ($SD = 7.3$) reported by Spielberger, Kling, and O'Hagen (1978) for 60 inmates anticipating an interview in a psychometric study, $t(106) = 4.33$, $p < .01$, one-tailed. In addition, the mean prestimulus heart rate level for our subjects (82.7 bpm) was substantially higher than the mean resting level (72.5 bpm) observed across six published studies by Hare and his colleagues in which heart rate data have been reported for inmate subjects (see Hare, 1986, for a review of these studies).

Although these data support the position that our subjects perceived the group contingency as a genuine threat and that they were concerned about their test outcomes, limits to external validity must be acknowledged. In real-life criminal investigations, the possible consequences of failing a polygraph test (e.g., arrest, prosecution, or incarceration) are likely to inspire even greater fear on the part of examinees. Although our testing situation was probably closer to real-life than the situations used in other mock crime studies, one must bear in mind that it was still an analogue situation.

Given the effectiveness of the threat manipulation, the results of our study strongly support the conclusion of Raskin and Hare (1978) that psychopaths are no more likely to "beat" a control question polygraph test than nonpsychopaths: 87% of our guilty subjects were correctly classified on the basis of blind analyses of the field polygraph charts, and the hit rates for psychopaths and nonpsychopaths were statistically equivalent. Out of the total sample of guilty subjects ($N = 24$), only 1 nonpsychopath and 2 psychopaths were classified as truthful based on the blind chart analyses, and one of the latter was correctly classified by the original examiner. Moreover, psychopaths and nonpsychopaths showed equal differentiation in their responses to relevant and control questions during the polygraph test, judging from the absence of any effects of psychopathy diagnosis on the magnitude of numerical scores. Even when extreme groups of psychopathic and nonpsychopathic subjects were selected according to their scores on an independent measure of socialization (the So scale of the CPI), differences failed to emerge.

Although contrary to clinical descriptions (e.g., Cleckley, 1976) that emphasize the equanimity with which psychopaths lie to others, these data are consistent with case reports suggesting that even the most hard-core psychopathic criminals can be detected when lying in a polygraph test situation (see, e.g., Cahill, 1986). It is possible that a lie detector test is too highly structured and restrictive a situation to provide psychopaths with a natural advantage over nonpsychopaths but that under less structured circumstances (e.g., extemporaneous lying, where perceived control is greater) these individuals might be able to lie with minimal physiological response. In addition, our findings do not rule out the possibility that sophisticated criminals trained in the deliberate use of physical or mental countermeasures might be able to escape detection. Honts, Hodes, and Raskin (1985) reported that 47% of their guilty subjects were able to beat a polygraph test after being trained to augment their responses to the control questions by biting their tongues or pressing their toes on the floor. Moreover, the polygraph examiner was unable to detect when subjects were using these countermeasures.

Our findings further suggest that in situations in which there is genuine concern about the outcome of a control question polygraph examination, a substantial proportion of innocent subjects may react more strongly to the relevant questions, producing deceptive test results. Within the group pressure atmosphere of this study, the accuracy of the CQT with innocent subjects was no better than chance (as compared with the 91% accuracy figure reported by Raskin & Hare, 1978). This finding is consistent with preliminary results from a large-scale field study recently reported (Patrick & Iacono, 1987), in which the

blind hit rate for confession-verified innocent subjects tested by experienced police examiners was only 55%. It is also consistent with the results of three other field studies of the CQT that collectively indicate about chance-level accuracy with innocent subjects (Barland & Raskin, 1976; Horvath, 1977; Kleinmuntz & Szucko, 1984). Although appreciable false positive error rates have been reported in other mock crime studies (e.g., Barland & Raskin, 1975; Forman & McCauley, 1986), and a link between the threat variable and false positive outcomes in our study cannot be demonstrated conclusively, some supportive evidence was obtained: A statistical analysis of false positive cases yielded results consistent with the hypothesis that subjects who failed the polygraph test adopted a defensive set during the test, in which higher levels of physiological arousal belied lower reports of state anxiety.[2] In contrast, differential hit rates were not associated with differences in examiner testing styles. Furthermore, our efforts to identify a unique response pattern for false positive versus true positive cases were entirely unsuccessful.

Statistical analyses of the numerical scores derived from the field polygraph charts revealed significant discrimination between guilty and innocent subjects for each of the three channels as well as for total scores. A discriminant analysis revealed that the SRR measure was the most important discriminator, followed by respiration. The independent contribution of the cardiovascular measure was minor. Based on a numerical analysis of charts derived from a laboratory polygraph, Raskin and Hare (1978) reported a similar pattern of results, with one important difference. In the earlier study, the discrimination between criterion groups was symmetric around zero, with the innocent subjects showing greater responsiveness to the control questions in each physiological channel, and the guilty subjects greater responsiveness to the relevant questions. The result was that the mean total numerical scores for the two groups, although differing in sign, were of approximately equal magnitude (innocent $M = +9.4$; guilty $M = -11.1$). In the present study, the discrimination was asymmetric, with guilty subjects producing a mean total of -14.33 and innocent subjects only $+1.43$. As a consequence, even an optimal placement of cutoff points failed to improve substantially the hit rate for innocent subjects.

The differences between our findings and those of Raskin and Hare (1978) suggest caution in generalizing the results of mock-crime laboratory polygraph studies to real-life. Although the change in motivational circumstances was the most noteworthy difference between the two studies, two other differences should be noted. The first is that in the earlier study the person responsible for administering and scoring the polygraph charts (Raskin) was also the primary investigator. This was not the case in our study. As a result, only in our investigation were the polygraph examiners blind to the base rate of guilt and innocence in the sample, as tends to be the case in real-life. Knowing that half of the charts he scored belonged to innocent subjects may have provided Raskin with an important decision-making advantage. A second point, emphasized by Lykken (1978, 1981), is that the charts were not independently rescored in the earlier study. This raises the possibility that subject behaviors and comments during the test may have contributed substantially to Raskin's hit rates.

In our study, there was little evidence to suggest that extrapolygraphic cues exerted a major influence on the original examiners' decisions. The coefficient of agreement between original and blind numerical scores was .85, and the original and blind scorers produced opposing conclusive decisions in only one case. Furthermore, there was no significant relation between examiner pretest guilt judgments and either (a) the guilt or innocence of the subject or (b) decisions based on the original numerical analyses. However, the results of this study certainly do not rule out a greater role for extrapolygraphic cues in the field testing situation. In a study in which police trainees underwent polygraph examinations under field-like circumstances, Ginton, Daie, Elaad, and Ben-Shakhar (1982) found that classifications based purely on behavioral observation were just as accurate as blind, chart-based decisions and that the two types of information together produced results superior to those based on either type of information alone. In addition, more recent research with the Royal Canadian Mounted Police in Canada (Patrick & Iacono, 1987) suggests that polygraph examiners' opinions may be influenced substantially by nonchart information in real-life cases. In our study, we were interested in the accuracy of the CQT as a psychophysiological test rather than the examiner's ability to extract the truth by other means (cf. Iacono & Patrick, 1987, 1988; Lykken, 1981). Therefore, in addition to having the charts rescored blindly, we made sure that our subjects were clear on how to deny involvement in the mock theft, and our examiners were asked not to press for a confession. As a result, the original examiners may not have had access to the same range of extrapolygraphic cues that tend to be available in a real testing situation.

[2] Of course, this finding may tell us little about what factors produce false positive outcomes in field polygraph situations. The variables mediating the impact of threatening consequences on real-life subjects are likely to vary with the nature of the sample and the nature of the threat. The important point is that outcomes for innocent subjects in this study were largely dependent on factors that had nothing to do with their criterion status.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Barland, G. H., & Raskin, D. C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology, 12,* 321–330.

Barland, G. H., & Raskin, D. C. (1976). *Validity and reliability of polygraph examinations of criminal suspects* (Report No. 76-1, Contract No. N1-99-0001). Washington, DC: National Institute of Justice, Department of Justice.

Cahill, T. (1986). *Buried dreams.* New York: Bantam Books.

Cleckley, H. (1976). *The mask of sanity* (5th ed.). St. Louis, MO: Mosby.

Clemmer, D. (1960). *The prison community.* New York: Holt, Rinehart & Winston.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Forman, R. F., & McCauley, C. (1986). Validity of the positive control question polygraph test using the field practice model. *Journal of Applied Psychology, 71,* 691–698.

Ginton, A., Daie, N., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology, 67,* 131–137.

Gough, G. H. (1969). *Manual for The California Psychology Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Hare, R. D. (1978). Electrodermal and cardiovascular correlates of psychopathy. In R. D. Hare & D. Schalling (Eds.), *Psychopathic behavior: Approaches to research* (pp. 107–143). Chichester, England: Wiley.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences, 1,* 111–119.

Hare, R. D. (1981). Psychopathy and violence. In J. R. Hays, T. K. Roberts, & K. S. Sloway (Eds.), *Violence and the violent individual* (pp. 53–74). New York: Spectrum.

Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 53,* 7–16.

Hare, R. D. (1986). Twenty years of experience with the Cleckley psychopath. In W. H. Reid, D. Dorr, J. I. Walker, & J. W. Bonner, III (Eds.), *Unmasking the psychopath* (pp. 3–27). New York: Norton.

Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology, 70,* 177–187.

Horvath, F. S. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology, 61,* 127–136.

Iacono, W. G., & Patrick, C. J. (1987). What psychologists should know about lie detection. In I. B. Weiner & A. Hess (Eds.), *Handbook of forensic psychology* (pp. 460–489). New York: Wiley.

Iacono, W. G., & Patrick, C. J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 205–233). New York: Guilford Press.

Kleinmuntz, B., & Szucko, J. J. (1984). A field study of the fallibility of polygraphic lie detection. *Nature, 308,* 449–450.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151–159.

Lykken, D. T. (1974). Psychology and the lie detector industry. *American Psychologist, 29,* 725–739.

Lykken, D. T. (1978). The psychopath and the lie detector. *Psychophysiology, 15,* 137–142.

Lykken, D. T. (1981). *A tremor in the blood: Uses and abuses of the lie detector.* New York: McGraw-Hill.

Patrick, C. J., & Iacono, W. G. (1987). Validity and reliability of the control question polygraph test: A scientific investigation [Abstract]. *Psychophysiology, 24,* 604–605.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15,* 126–136.

Reid, J. E., & Inbau, F. E. (1977). *Truth and deception: The polygraph ("lie detector") technique* (3rd ed.). Baltimore, MD: Williams & Wilkens.

Schmauk, F. J. (1970). Punishment, arousal, and avoidance learning in sociopaths. *Journal of Abnormal Psychology, 76,* 325–335.

Schroeder, M. L., Schroeder, K. G., & Hare, R. D. (1983). Generalizability of a checklist for assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 51,* 511–516.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Kling, J. K., & O'Hagen, S. E. J. (1978). Dimensions of psychopathic personality: Antisocial behavior and anxiety. In R. D. Hare & D. Schalling (Eds.), *Psychopathic behavior: Approaches to research* (pp. 23–46). Chichester, England: Wiley.

Sykes, G. M. (1958). *The society of captives: A study of a maximum security prison.* New York: Atheneum.

Thomas, C. W., & Peterson, D. M. (1977). *Prison organization and inmate subcultures.* Indianapolis, IN: Bobbs-Merrill.

Waid, W. M., Orne, M. T., & Wilson, S. K. (1979). Effects of level of socialization on electrodermal detection of deception. *Psychophysiology, 16,* 15–22.

Received February 16, 1988
Revision received July 26, 1988
Accepted August 9, 1988 ■